

بررسی میزان اثربخشی عناصر ابرداده ای بر رتبه بندی صفحات وب توسط موتورهای کاوش عمومی

چکیده:

پژوهش حاضر با هدف تعیین میزان اثربخشی عناصر ابرداده ای بر رتبه بندی صفحات وب توسط سه موتور کاوش عمومی گوگل، یاهو، و ام اس ان انجام پذیرفت. اثربخشی پنج عنصر ابرداده ای نشانه عنوان زبان اچ تی ام ال، ابرنشانه های کلیدواژه ها و توصیف زبان اچ تی ام ال، ابرداده های عنوان، و موضوع از قالب ابرداده ای دابلین کور، که بر بازنمون موضوعی صفحات وب متمرکزند با روش تجربی مورد آزمون قرار گرفت. ۸۴ صفحه وب در گروه های گواه و آزمون به عنوان جامعه مورد مطالعه تحت دامنه فرعی <http://metadata.irandoc.ac.ir> منتشر شد. از میان موتورهای کاوش انتخابی دو موتور کاوش یاهو و گوگل صفحات را نمایه سازی نمودند. کلیدواژه های منحصر به فرد و تکرار شونده که در مرحله طراحی در قالب عناصر ابرداده ای به صفحات گروه آزمون افزوده شده بود مورد کاوش قرار گرفت و معنی داری تفاوت میان رتبه صفحات دو گروه با استفاده از آزمون ناپارامتری یو من وایتنی محاسبه گردید. نتایج نشانگر اثربخشی دو عنصر عنوان و توصیف زبان اچ تی ام ال در هر دو موتور کاوش مورد مطالعه است. کلیدواژه های منحصر به فرد در ابرنشانه کلیدواژه های اچ تی ام ال و ابرداده موضوع دابلین کور به بهبود رتبه صفحات در یاهو انجامید. در حالی که تکرار کلیدواژه ها در بخش سرآیند صفحه وب، بهبود رتبه را در موتور کاوش یاهو سبب گشته است، گوگل عملکردی معکوس نشان می دهد.

کلیدواژه ها: ابرداده، ابرنشانه های اچ تی ام ال، دابلین کور، بازیابی اطلاعات، رتبه بندی نتایج کاوش، موتورهای

کاوش عمومی، وب، دسترس پذیری صفحات وب

عاطفه شریف^۱

۱. مقدمه

همزمان با تولد وب جهان شاهد تغییرات شگرفی در حوزه ذخیره و بازیابی اطلاعات بوده است. هر روز بر تعداد صفحات وب افزوده می گردد و حجم وسیعی از اطلاعات، در بستر وب، به صورت ساختار نیافته^۲ (یالتاغان^۳، ۲۰۰۲) و فارغ از کنترل محتوایی و کتابشناختی منتشر می شود (زانگ و جاستریم^۴، ۲۰۰۴؛ گوتلیب و الیوپولوس^۵، ۲۰۰۳). در چنین وضعیتی مسئله اساسی چگونگی کنترل و مدیریت بدنه ساختار نیافته و رشد شتابناک این بدنه است (اسدی و جمالی مهمویی^۶، ۲۰۰۴). تاکنون ابزارهای کاوش^۷ - از قبیل موتورهای کاوش^۸ ابرموتورهای کاوش^۱، راهنماهای موضوعی^۲، نرم افزارهای کاوش^۳ - محیط مجازی وب را تا اندازه ای تحت کنترل و مدیریت خویش درآورده اند.

^۱ دانشجوی دکتری علوم کتابداری و اطلاع رسانی دانشگاه فردوسی مشهد e-mail:atefehsharif@gmail.com

^۲ Unstructured

^۳ Yaltaghian

^۴ Zhang & Jastram

^۵ Gotlieb & Eliopoulos

^۶ Asadi & Jamali M.

^۷ Internet search tools

^۸ Search engines

کاربران از میان ابزارهای کاوش رایج، موتورهای کاوش را به عنوان نقطه آغازین ورود به اینترنت تلقی می کنند (اسپینک و دیگران^۴، ۲۰۰۱ نقل در دوال و واگان^۵، ۲۰۰۴؛ بار-ایلان^۶، ۲۰۰۵؛ زانگ و دیمیتروف^۷، ۲۰۰۴) بیش از ۹۵ درصد ترافیک کاوش در اینترنت مربوط به موتورهای کاوش است و ۸۰ درصد کاربران، اطلاعات مورد نیاز خود را از طریق موتورهای کاوش به دست می آورند (هاتلی^۸، ۲۰۰۲ نقل در زانگ و دیمیتروف^۹، ۲۰۰۵). یافتن اطلاعات موضوعی ویژه در وب با دشواری هایی همراه است و هر روز بر این دشواری افزوده می گردد (دروت^۹، ۲۰۰۰، ص ۲۰۹). تعداد نتایج بازیابی شده توسط موتورهای کاوش اغلب چنان فراوان است که کاربر عملاً جز مرور چند صفحه نخست نتایج، از سایر صفحات منصرف می شود (جانسن، اسپینک، و ساراسویک^{۱۰}، ۲۰۰۰؛ فدایی عراقی^{۱۱}، ۲۰۰۵، ص ۱۳؛ یالتاقیان، ۲۰۰۲) و به ناچار بر رتبه بندی^{۱۲} ارائه شده از سوی موتورهای کاوش اعتماد می کند (بار-ایلان، ۲۰۰۵). در این وضعیت، چنانچه صفحه ای مرتبط، در رتبه های اول جای نگیرد ممکن است از دید کاوشگر پنهان بماند (زانگ و جاستریم، ۲۰۰۵، ص ۹۲؛ گوتلیب و الیوپولوس، ۲۰۰۳).

از سویی، یکی از اولین دغدغه های ناشران وب سایت ها، دستیابی به رتبه های برتر در میان وب سایت های مشابه و هم موضوع است. به این منظور همواره سیاهه ای از عناوین^{۱۳}، کلیدواژه ها^{۱۴} و توصیف هایی^{۱۵} که احتمال کسب رتبه های برتر را به دنبال دارند، تهیه و در طراحی صفحات لحاظ می شود (ریچاردسون^{۱۶}، ۲۰۰۳ نقل در زانگ و دیمیتروف^{۱۷}، ۲۰۰۵a). طراحی نرم افزارهایی چون تحلیلگر چگالی کلیدواژه^{۱۷}، ورد ترکر^{۱۸}، وب پزیشن گولدر^{۱۹}، و شکل گیری و گسترش وب سایت هایی که خدمات توصیه ای و مشاوره ای بهینه سازی صفحات^{۲۰} را به منظور کسب رتبه های برتر در موتورهای کاوش ارائه می دهند، تأییدی بر حساسیت و توجه بر این مسئله است.

نتایج تحقیقات، تفاوت های قابل ملاحظه ای را میان الگوریتم های رتبه بندی موتورهای کاوش عمومی نشان می دهد (بار-ایلان، ۲۰۰۵). اطلاعات مربوط به الگوریتم های رتبه بندی موتورهای کاوش به صورت طبقه بندی شده^{۲۱} و

¹ Meta search engines

² Directories

³ Search Utilities

⁴ Spink et al

⁵ Thelwall & Vaughan

⁶ Bar_Ilan

⁷ Zhang & Dimitroff

⁸ Haltley

⁹ Drott

¹⁰ Jansen ; Spink ; and Saracevic

¹¹ Fadaie Araghi

¹² Ranking

¹³ Title

¹⁴ Keyword

¹⁵ Description

¹⁶ Richardson

¹⁷ Keyword Density analyzer

¹⁸ word tracker

¹⁹ web poisons gold

²⁰ Search Engine Optimization (SEO)

²¹ Classified

به عنوان اسرار تجاری^۱ محافظت می شود. حفظ حالت رقابتی و جلوگیری از سوء استفاده طراحان وب سایتها از این اطلاعات، از جمله دلایل حفاظت به شمار می آیند (بار-ایلان، ۲۰۰۵، ص ۱۵۱۲). با وجود ابهام های موجود در زمینه الگوریتم رتبه بندی موتورهای کاوش، پژوهش پیرامون چگونگی رتبه بندی نتایج متوقف نشده است و تلاشهایی در زمینه کشف عوامل اثرگذار و تعیین میزان اثرگذاری هر یک بر رتبه بندی صفحات انجام پذیرفته است (ترنر و برک بیل^۲، ۱۹۹۸؛ زانگ و دیمیتروف، ۲۰۰۴؛ ۲۰۰۵a، ۲۰۰۵b، صفری^۳، ۲۰۰۵؛ محمد^۴، ۲۰۰۶؛ هنشا و والاسکاس^۵، ۲۰۰۱).

۲. بیان مسئله

صفحات وب متشکل از سه جزء اند: ۱. معنا^۶ یا محتوا^۷ ۲. بستر نحوی^۸ یا ساختار^۹ و ۳. پیوندهای فرامتنی^{۱۰}. محتوا در بستر نحوی زبانهای نشانه گذاری^{۱۱} که قالب یا ساختار ارائه محتوا را فراهم می آورند جای می گیرد و ارتباط میان اجزای اطلاعاتی از طریق پیوندهای فرامتنی حاصل می شود. هر یک از این عناصر - محتوا، ساختار و پیوندهای فرامتنی - ویژگی های خاصی دارند که بستر ارزیابی صفحات را فراهم می آورد و به طور بالقوه در بهبود کیفی رتبه بندی حاصل از کاوش اثر گذار است (کوودو-تررو^{۱۲}، ۲۰۰۴).

گوگل بیش از هزار عامل را در نظام رتبه بندی نتایج خود مدنظر دارد؛ اما به دلیل ماهیت تجاری و حفظ یکپارچگی نتایج کاوش، از ذکر جزئیات بیشتر خودداری می کند (گوگل^{۱۳}، ۲۰۰۴). توافقی بر سر موثرتر بودن عاملی نسبت به سایر عوامل وجود ندارد (فیشکین^{۱۴}، ۲۰۰۵) و هر یک از موتورهای کاوش، الگوریتم رتبه بندی خاصی را دنبال می کنند (هنشا و والاسکاس، ۲۰۰۱، ص ۹۲). ساختار ابر داده ای^{۱۵}، محتوای صفحه^{۱۶} - عوامل داخلی - و وضعیت ارجاعات فرامتنی^{۱۷} - عوامل خارجی - از جمله عوامل موثر بر رتبه بندی ذکر شده است (زانگ و جاستریم، ۲۰۰۵).

در این پژوهش، از میان عوامل مؤثر شناخته شده بر رتبه بندی نتایج کاوش، تمرکز بر عناصر ابر داده ای^{۱۸} است، و از میان عناصر ابر داده ای، پنج عنصر، به عنوان معیار ربط فنی^{۱۹} انتخاب شده و میزان اثرگذاری هر یک - به صورت

¹ Trade Secrets

² Truner & Brackbill

³ Safari

⁴ Mohamed

⁵ Henshaw & Valauskas

⁶ Semantic

⁷ Content

⁸ Syntax

⁹ Structure

¹⁰ Hyper- Links

¹¹ Markup Languages

¹² Quevedo_Torrero

¹³ google

¹⁴ Fishkin

¹⁵ Metadata Structure

¹⁶ Webpage Content

¹⁷ Hyperlink Cited status

¹⁸ Metadata elements

¹⁹ Technical Relevance

منحصر به فرد و در صورت تکرار - بر رتبه صفحات بازیابی شده توسط موتورهای کاوش عمومی مورد بررسی و تجزیه و تحلیل قرار گرفته است. عناصر ابرداده ای مورد بررسی در این پژوهش عبارتند از: نشانه عنوان^۱ از زبان اچ تی ام ال، ابرنشانه های کلیدواژه ها^۲ و توصیف^۳ از زبان اچ تی ام ال، دو ابرداده عنوان^۴ و موضوع^۵ از قالب ابرداده ای دابلین کور

اگر چه نشانه عنوان زبان اچ تی ام ال، عنصر ابرداده ای محسوب نمی شود؛ اما به واسطه اهمیت ویژه اش (سالیوان^۶، ۲۰۰۲؛ نوروزی^۷، ۲۰۰۵)، در میان سایر نشانه های اچ تی ام ال، در کنار چهار عنصر ابرداده ای دیگر مورد بررسی قرار گرفته است.

پنج عنصر منتخب، تنها تعدادی از عناصر ابرداده ای موجود در قالب ابرداده ای دابلین کور و زبان اچ تی ام ال می باشند. این عناصر بر بازنمون موضوعی مدرک متمرکزند و در برخی منابع کاربرد آنها توصیه شده است (زانگ و دیمتروف، ۲۰۰۴؛ سالیوان، ۲۰۰۲؛ لی - اسملتزر^۸، ۲۰۰۰، ص ۲۰۶؛ والکی، فریر^۹، ۲۰۰۱، ص ۲۷۲).

۳. پرسش های اساسی

پژوهش حاضر فاقد فرضیه، و پرسش های اساسی آن به این قرار است:

۱. حضور کلیدواژه مورد کاوش در هر یک از عناصر مورد بررسی (نشانه عنوان، ابرنشانه های کلیدواژه ها و توصیف زبان نشانه گذاری فرامتن، ابرداده های عنوان و موضوع قالب ابرداده ای دابلین کور) تا چه میزان بر رتبه بندی صفحات در سه موتور کاوش مورد آزمون اثرگذار است؟
۲. تکرار کلیدواژه مورد کاوش در دو تا پنج عنصر یاد شده در بخش سرآیند^{۱۰} صفحه و صرفنظر از نوع عنصر، تا چه میزان بر بهبود رتبه صفحه اثرگذار است؟

۴. روش شناسی

پژوهش هایی که به تعیین میزان اثربخشی عناصر ابرداده ای بر رتبه بندی صفحات وب در موتورهای کاوش پرداخته اند با تفاوت هایی اندک از روش تجربی بهره برده اند. در این پژوهش نیز با آگاهی از وجود سایر عواملی که بر رتبه بندی صفحات اثرگذارند، جهت اعمال متغیرهای مستقل، کنترل متغیرهای دخیل، و مشاهده تغییرات در متغیر وابسته از روش تجربی استفاده شده است.

^۱<title> </title>

^۲<meta name="description" content=" ">

^۳<meta name="keywords" content=" ">

^۴<meta name="DC.Title" content=" ">

^۵<meta name="DC.Subject" content=" ">

^۶ Salivan

^۷ Noruzi

^۸ Lee-Smeltzer

^۹ Valqui & Freire

^{۱۰} Header

جامعه مورد پژوهش، نشریه ای الکترونیکی با ۸۴ صفحه وب، طراحی شده توسط پژوهشگر است. این نشریه با روندی که در ادامه می آید طراحی شد و سپس در معرض نمایه سازی موتورهای کاوش قرار گرفت.

۴-۱. مراحل طراحی صفحات

در مرحله نخست تعداد چهارده مقاله در موضوعات وب معنایی^۱، هستی شناسی^۲، وب سنجی^۳ و ابر داده های دابلین کور از میان مقالات منتشر شده در نشریات رایگان حوزه کتابداری و اطلاع رسانی که سیاهه این نشریات در فهرست راهنمای نشریات پیوسته رایگان^۴ معرفی شده است، انتخاب گردید.

کلیدواژه های نمایه ای این مقالات به شیوه کنترل نشده - زبان طبیعی - و از بستر عنوان، کلیدواژه ها (در صورت وجود) چکیده، و در برخی موارد متن مقاله برگزیده شد. پژوهش بر دو دسته کلیدواژه متمرکز است؛ الف. کلیدواژه های منحصر به فرد؛ و ب. کلیدواژه های تکرار شونده. کلیدواژه های منحصر به فرد آن دسته از کلیدواژه هایی هستند که منحصرأ به یک عنصر ابر داده ای اختصاص یافته است، با دیگر کلیدواژه های مربوط به سایر عناصر همپوشانی و شباهت ندارد، و پاسخگویی به نخستین پرسش را ممکن می سازد. کلیدواژه های تکرار شونده آن دسته از کلیدواژه هایی هستند که در تمام عناصر ابر داده ای به یک شیوه و ترتیب، قبل از کلیدواژه های منحصر به فرد، و به منظور سنجش اثربخشی تکرار کلیدواژه ها در دو تا پنج عنصر ابر داده ای - پرسش دوم - افزوده شده اند.

به طور نمونه کلیدواژه های منحصر به فرد و تکرار شونده مقاله هشتم با عنوان **A Metadata Registry for the Semantic Web** در جدول ۱ آمده است.

جدول ۱. نمونه ای از کلیدواژه های منحصر به فرد و تکرار شونده

Element	Keywords	
	Unique	Repeated
E1 ⁵	---	Semantic web, Metadata Registry
E2	Metadata schemas, DCMI vocabulary	Semantic web, Metadata Registry
E3	Automated sharing of metadata, Dublin core metadata Initiative, DCMI Registry	Semantic web, Metadata Registry
E4	---	Semantic web, Metadata Registry
E5	Registry applications	Semantic web, Metadata Registry

پس از تعیین کلیدواژه های مربوط به هر یک از عناصر مورد بررسی، در مرحله بعد، با توجه به پرسش های اساسی پژوهش گروه های گواه و آزمون تعیین گردید. پاسخگویی به پرسش ها با وجود شش گروه صفحه امکان پذیر

¹ Semantic web

² Ontology

³ Webometrics

⁴ Directory of Open Access Journals , Available at: www.doaj.org

⁵ عنصر اول: عنوان اچ تی ام ال

است. گروه بندی صفحات در جدول ۲ آمده است. به منظور حفظ هم ترازى در دو گروه گواه و آزمون تعداد صفحات گروه گواه با تعداد صفحات گروه آزمون برابر است، و محتوای صفحه آزمون در هر مقاله جز در مورد عناصر ابر داده ای افزوده شده مشابه با صفحه گروه گواه خود است. به این ترتیب چهارده مقاله منتخب در شش شماره نشریه تکرار شده اند (۸۴ = ۱۴ * ۶).

جدول ۲. گروه بندی صفحات

گروه	حاوی کلیدواژه منحصر به فرد در	حاوی تکرار کلیدواژه ها در
اول	----	----
دوم	----	نشانه عنوان اچ تی ام ال (عنصر ۱)
سوم	ابرنشانه کلیدواژه های اچ تی ام ال (عنصر ۲)	نشانه عنوان و ابرنشانه کلیدواژه های اچ تی ام ال (عناصر ۱، ۲ و ۳)
چهارم	ابرنشانه توصیف اچ تی ام ال (عنصر ۳)	نشانه عنوان، ابرنشانه های کلیدواژه ها و توصیف اچ تی ام ال (عناصر ۱، ۲، ۳ و ۴)
پنجم	----	نشانه عنوان، ابرنشانه های کلیدواژه ها و توصیف اچ تی ام ال، ابر داده عنوان دابلین کور (عناصر ۱، ۲، ۳، ۴ و ۵)
ششم	ابرداده موضوع قالب ابر داده ای دابلین کور (عنصر ۵)	نشانه عنوان، ابرنشانه های کلیدواژه ها و توصیف اچ تی ام ال، ابر داده های عنوان و موضوع دابلین کور (عناصر ۱، ۲، ۳، ۴، ۵ و ۶)

سپس کلیدواژه های منتخب در بستر نحوی اچ تی ام ال نشانه گذاری و در کد منبع^۱ مقالات گنجانده شد. به طور نمونه کدگذاری مقاله هشتم چنین است: در این مثال کلیدواژه های تکرار شونده به صورت توپر نمایش داده شده است. گروه نخست فاقد کلیدواژه است.

در گروه دوم

```
<TITLE>A Metadata Registry for the Semantic Web</TITLE>
```

در گروه سوم

```
<TITLE>A Metadata Registry for the Semantic Web</TITLE>
<META name="keywords" content="Semantic web, Metadata Registry, Metadata schemas, DCMI
vocabulary">
```

در گروه چهارم

```
<TITLE>A Metadata Registry for the Semantic Web</TITLE>
<META name="keywords" content="Semantic web, Metadata Registry, Metadata schemas, DCMI
vocabulary">
<META name="description" content="Semantic web, Metadata Registry, Automated sharing of metadata,
Dublin core metadata Initiative, DCMI Registry">
```

^۱ Source code

```
<TITLE>A Metadata Registry for the Semantic Web</TITLE>
<META name="keywords" content="Semantic web, Metadata Registry, Metadata schemas, DCMI
vocabulary">
<META name="description" content="Semantic web, Metadata Registry, Automated sharing of metadata,
Dublin core metadata Initiative, DCMI Registry">
<META name="DC.Title" content="A Metadata Registry for the Semantic Web">
```

```
<TITLE>A Metadata Registry for the Semantic Web</TITLE>
<META name="keywords" content="Semantic web, Metadata Registry, Metadata schemas, DCMI
vocabulary">
<META name="description" content="Semantic web, Metadata Registry, Automated sharing of metadata,
Dublin core metadata Initiative, DCMI Registry">
<META name="DC.Title" content="A Metadata Registry for the Semantic Web">
<META name="DC.Subject" content="Semantic web, Metadata Registry, Registry applications">
```

۴-۲. انتشار صفحات طراحی شده

به این ترتیب ۸۴ صفحه ایستا^۱ به کمک واژه پرداز ورد از مجموعه مایکروسافت آفیس ۲۰۰۳^۲ طراحی و از ۲۴ خرداد تا ۸ شهریور ۱۳۸۵^۳ در دامنه فرعی "ابرداده" از سایت پژوهشگاه اطلاعات و مدارک علمی ایران^۴ منتشر شد، و در معرض نمایه سازی سه موتور کاوش گوگل^۵، یاهو^۶، ام اس ان^۷ - سه موتور کاوش پر استفاده به استناد الکسا^۸، سرچ اینجین و اچ^۹، سرچ اینجین گاید^{۱۰} و سایت های مشابه^{۱۱} - قرار گرفت. در دهمین روز انتشار صفحات، موتور کاوش گوگل، تمامی صفحات منتشر شده را نمایه سازی نمود؛ این در حالی است که یاهو روندی کندتر داشت و پس از گذشت بازه زمانی مشخص شده، ۷۶ صفحه مربوط به مقالات نشریه نمایه شد. اما از میان سه موتور کاوش انتخابی، ام اس ان به دلیل عدم نمایه سازی تعداد قابل قبولی از صفحات در مدت زمان تعیین شده، از مجموعه پژوهش حذف شد، و ادامه پژوهش با تمرکز بر صفحات نمایه شده در دو موتور کاوش یاهو و گوگل انجام پذیرفت.

¹ Static

² Microsoft Office Word 2003

³ June 14, 2006 till Aug 30, 2006

⁴ <http://metadata.irandoc.ac.ir>

⁵ www.google.com

⁶ www.yahoo.com

⁷ www.msn.com

⁸ Jun, 11, 2006 From: http://www.alexa.com/site/ds/top_500

⁹ www.searchenginewatch.com

¹⁰ www.searchengineguid.com

¹¹ <http://www.seoconsultants.com/search-engines/>

۴-۳. گردآوری داده ها

به منظور تعیین میزان اثربخشی عناصر ابرداده ای در این مرحله از پژوهش، کاوش کلیدواژه ای در کادر محاوره ای ساده^۱ دو موتور کاوش گوگل و یاهو انجام پذیرفت. از آنجایی که دو گروه گواه و آزمون، در وب سایت طراحی شده گنجانده شده است؛ و نیازی به تعیین رتبه صفحات در میان سایر صفحات موجود در وب نیست، کاوش کلیدواژه ای در دامنه وب سایت طراحی شده محدود شد. فرمول کاوش در موتور کاوش گوگل و یاهو به ترتیب چنین است:

Site:<http://metadata.irandoc.ac.ir> "Keywords"

Site:metadata.irandoc.ac.ir "Keywords" یا Domain:metadata.irandoc.ac.ir "Keywords"

به منظور تعیین میزان اثربخشی هر یک از عناصر ابرداده ای (پرسش نخست پژوهش) دو مجرای آزمون وجود دارد:

- کاوش کلیدواژه های منحصر به فرد که تنها به عنصر ابرداده ای ویژه ای اختصاص یافته است و مقایسه رتبه صفحه حاوی آن کلیدواژه با صفحه گروه گواه مورد نظر؛
- کاوش کلیدواژه های تکرار شونده؛ چنان که بیان شد به ترتیب افزایش شماره های نشریه، یک عنصر ابرداده ای به عناصر قبلی افزوده می گردد و تنها تفاوت میان دو مقاله مشابه در دو نشریه با شماره های متوالی، حضور کلیدواژه های تکرار شونده و منحصر به فردی است که در عنصر ابرداده ای بعدی افزوده شده است. لذا هر یک از شماره های نشریه در حالی که حاوی مقالات گروه آزمون نشریه شماره قبل خود است، حاوی مقالات گروه گواه نشریه شماره بعد نیز می باشد.

به این ترتیب به استثنای عناصر عنوانی (عنوان اچ تی ام ال و عنوان دابلین کور) دو دسته آزمون به منظور تعیین اثربخشی حضور هر یک از عناصر ابرداده ای (پرسش نخست) قابل انجام است.

رتبه حاصل از هر یک از کاوش های انجام شده در موتورهای کاوش، به تفکیک موتور کاوش و گروه گواه و آزمون با توجه به سوالات پژوهش در نرم افزار آماری SPSS^۲ وارد شد و مورد تجزیه و تحلیل قرار گرفت. با توجه به این که نتایج حاصل از کاوش در موتورهای کاوش از نوع داده های رتبه ای است (ترنر و برک بیل، ۱۹۹۸؛ صفری^۳، ۲۰۰۵)، میزان معنی دار بودن تفاوت میان دو گروه گواه و آزمون - دو گروه مستقل - در بازایی صفحات وب، با استفاده از آزمون ناپارامتری یو من وایتنی^۴ تعیین گردید.

در مجموع، تحت شرایط زمانی و مکانی یکسان، ۳۶۳ کاوش در موتور کاوش گوگل و ۲۵۹ کاوش در موتور کاوش یاهو انجام، رتبه های کسب شده در دو گروه صفحات آزمون و گواه ثبت شد و تجزیه و تحلیل آماری بر مبنای داده های جمع آوری شده، توسط نرم افزار آماری اس پی اس انجام پذیرفت.

۵. تجزیه و تحلیل یافته ها

^۱ Simple Search Box

^۲ Statistical Package for Social Science

^۳ Safari

^۴ Mann-Whitney U

۵-۱. پاسخگویی به نخستین پرسش پژوهش: میزان اثربخشی هر یک از عناصر ابرداده ای بر رتبه بندی توسط دو موتور کاوش یاهو و گوگل

به منظور درک بهتر و امکان مقایسه و تحلیل نتایج، جدول ۳ که نمایی کلی از مقدار پی محاسبه شده توسط نرم افزار آماری اس پی اس در آزمون یو من وایتنی است، ارائه می شود. وجود تفاوت معنی دار در هر عنصر، با قلم^۱ درشت تر و توپر نشان داده شده است. تفاوت میان میانگین رتبه های کسب شده در هر گروه نشانگر آن است که رتبه به نفع کدامین گروه تغییر وضعیت داده است؛ لذا در هر مورد از عناصر، به تفکیک موتور کاوش، گروه صفحاتی که بهبود رتبه داشته - آزمون یا گواه- با اصطلاح "به نفع" مشخص شده است.

جدول ۳. نتایج آزمون یو من وایتنی در پاسخ به پرسش نخست

عناصر مورد آزمون		نوع کلیدواژه		گوگل		ياهو	
				Asymp. Sig. (2-tailed)		Asymp. Sig. (2-tailed)	
				به نفع		به نفع	
عنوان اچ تی ام ال	کلیدواژه منحصر به فرد	---		---		---	
	کلیدواژه تکرار شونده	.000		.001		آزمون	
کلیدواژه های اچ تی ام ال	کلیدواژه منحصر به فرد	.462		.017		آزمون	
	کلیدواژه تکرار شونده	.472		.172		آزمون	
توصیف اچ تی ام ال	کلیدواژه منحصر به فرد	.015		.000		آزمون	
	کلیدواژه تکرار شونده	.016		.000		گواه	
عنوان دابلین کور	کلیدواژه منحصر به فرد	---		---		---	
	کلیدواژه تکرار شونده	.826		.807		آزمون	
موضوع دابلین کور	کلیدواژه منحصر به فرد	.110		.023		آزمون	
	کلیدواژه تکرار شونده	.038		.050		گواه	

یافته های حاصل از بررسی اثربخشی دو ابرنشانه کلیدواژه ها و توصیف در این پژوهش، با یافته های ترنر و برک بیل (۱۹۹۸) مبنی بر اثربخشی ابرنشانه کلیدواژه ها در مقایسه با ابرنشانه توصیف متفاوت است. یافته ها در پژوهش حاضر نشان داد که ابرنشانه توصیف در مقایسه با ابرنشانه کلیدواژه ها اثربخشی بیشتری در بهبود رتبه صفحات دارد.. یافته های صفری (۲۰۰۵) مبنی بر عدم اثربخشی عنصر عنوان دابلین کور بر رتبه صفحات در این پژوهش نیز تأیید می شود. یافته های این پژوهش در راستای نتایج زانگ و دیمیتروف (۲۰۰۴) اثربخشی عنصر موضوع دابلین کور در بهبود رتبه صفحات را نشان می دهد. این یافته با یافته های صفری مبنی بر عدم اثربخشی عنصر موضوع دابلین کور متفاوت است.

¹ font

۵-۲. پاسخگویی به پرسش دوم: میزان اثربخشی دو تا پنج بار تکرار کلیدواژه در بخش سرآیند صفحات وب بر رتبه بندی صفحات وب در دو موتور کاوش یاهو و گوگل

جدول شماره ۴، نمایی کلی از مقدار پی محاسبه شده توسط نرم افزار آماری اس پی اس در آزمون یو من وایتنی ارائه می کند. در پاسخ به پرسش دوم پژوهش، اثربخشی حضور کلیدواژه های تکرار شونده در دو تا پنج عنصر تحت بررسی مورد تجزیه و تحلیل قرار گرفت.

جدول ۴. نتایج آزمون یو من وایتنی در پاسخ به پرسش دوم

تعداد تکرار مورد بررسی		گوگل		ياهو	
		به نفع	Asymp. Sig. (2-tailed)	به نفع	Asymp. Sig. (2-tailed)
۲ بار تکرار		آزمون	.580	آزمون	.172
۳ بار تکرار		گواه	.008	آزمون	.016
۴ بار تکرار		گواه	.062	آزمون	.019
۵ بار تکرار		گواه	.000	آزمون	.175

عملکرد دو موتور کاوش گوگل و یاهو در مقابل عناصر ابرداده ای در این پژوهش مورد بررسی قرار گرفت. در بیانی کوتاه عملکرد دو موتور کاوش مورد بررسی چنین است:

- موتورهای کاوش نسبت به حضور عناصر ابرداده ای مورد بررسی بی اعتنا نیستند؛
- عنصر عنوان اچ تی ام ال در مقایسه با سایر عناصر مورد بررسی، اثرگذارترین عنصر در هر دو موتور کاوش مورد بررسی است؛
- از میان ابرنشانه های اچ تی ام ال، ابرنشانه توصیف اثربخشی بیشتری در مقایسه با ابرنشانه کلیدواژه ها دارد؛
- توجه به عناصر ابرداده ای زبان اچ تی ام ال بیش از عناصر ابرداده ای قالب دابلین کور است؛
- توجه موتورهای کاوش مورد بررسی به عنصر موضوع قالب ابرداده ای دابلین کور بیش از عنصر عنوان آن است؛
- دو موتور کاوش گوگل و یاهو، عملکرد متفاوتی نسبت به کلیدواژه های تکرار شونده و منحصر به فرد دارند؛
- در هر دو موتور کاوش، کلیدواژه های منحصر به فرد در مقایسه با کلیدواژه های تکرار شونده، اثربخشی بیشتری دارد؛
- موتور کاوش گوگل بیش از یاهو نسبت به کلیدواژه های تکراری در بخش سرآیند صفحات وب حساس است و به نظر می رسد که این تکرارها را نوعی تقلب به شمار می آورد.

۶. نتیجه گیری

با وجود تمامی پیشرفت های حاصل شده در فناوری موتورهای کاوش، هنوز روش اصلی نمایه سازی در موتورهای کاوش، محتوا محور است و بر مبنای کلیدواژه ها انجام می پذیرد (یو و دیگران^۱، ۲۰۰۲). نمایه سازی تمام متن، کاوش کلیدواژه ای بدون توجه به بافت متن و کاستی های موجود در رتبه بندی ها منجر به جامعیت بالا و مانعیت پایین می شود (لی - اسملتزر، ۲۰۰۰، ص ۲۰۶). تحقیقات فراوانی در زمینه مشکلات نمایه سازی خودکار در موتورهای کاوش انجام گرفته است (چانگ و لی^۲، ۱۹۹۸، ص ۱۴۹). آنچه پژوهشگران در باب آن به اتفاق نظر رسیده اند آن است که نمایه سازی غیرمتمرکز و توزیع شده، به گونه ای که در موتورهای کاوش کنونی در جریان است؛ معماری غیر قابل تاییدی در بازیابی اطلاعات وب است (لی - اسملتزر، ۲۰۰۰، ص ۲۰۶).

چنانچه محتوای صفحه به درستی در فیلد های ابرداده ای بازنمون گردد و چنانچه موتورهای کاوش از این فیلدهای ابرداده ای در رتبه بندی و بازیابی نتایج بهره برند؛ مانعیت افزایش و نتایج بازیابی به سطحی قابل مدیریت کاهش می یابد. به علاوه رتبه بندی صفحات از صحت بیشتری برخوردار خواهد شد (زانگ و جاستریم، ۲۰۰۵، ص ۱) و آن بخش از ویژگی هایی که شناسایی آن به درک انسانی نیاز دارد، در فرایند نمایه سازی منظور خواهد شد. اما همچنان پاسخ به این سوال که آیا جامعه بهره گیر به اهمیت وجود ابرداده پی برده است یا خیر، آن را تایید و یا به درستی از آن بهره می برد، یک معماست (زانگ و جاستریم، ۲۰۰۵، ص ۲۲).

در دیدگاه تیم برنرزی^۳ و سایر پیشگامان وب معنایی، وب کنونی اطلاعاتی انسان فهم دارد. صفحات وب به زبان اچ تی ام ال نگاشته می شود و مرورگرها بر نشانه هایی تمرکز دارند که نه بر محتوای معنایی بلکه بر ساختار متمرکزند. چنانچه وب معنایی تحقق یابد صفحات وب آینده همان گونه که انسان فهم هستند، ماشین فهم^۴ نیز خواهند بود (کمپل^۵، ۲۰۰۴، ص ۱۹۷). به این گونه که اطلاعات از طریق نشانه های معنادار ارائه می شود و ابرداده های لازم توسط آر دی اف به صفحات وب افزوده می شود. پدیدآورندگان وب فردا، امکان جست جوپذیر کردن سایت های خود را با مانعیت بیشتری عملی خواهند ساخت و کارگزاران هوشمند^۶ آینده، بسیار قدرتمندتر و متکامل تر از موتورهای کاوش امروزی عمل خواهند کرد (کمپل، ۲۰۰۴، ص ۱۹۹).

¹ Yu et al

² Chung & Lee

³ Tim Burners Lee

⁴ Machine understandable

⁵ Campbell

⁶ Intelligent Agents

- Asadi, S., & Jamali M., H.R. (2004). "Shifts in search engine development: A review of past, present and future trends in research on search engines". *Webology*, 1(2). Retrieved Feb, 18, 2005, From <http://www.webology.ir/2004/v1n2/a6.html>
- Bar_Ilan, J. (2005). " Comparing Rankings of Search Results on the Web" . *Information Processing & Management* , 41. Retrieved Oct, 12, 2005, From Elsevier Database.
- Campbell, D.G. (2004). "The Metadata – Bibliographic Organization nexus" In *Metadata Applications and Management*. London: Facet Publishing, 185-203.
- Chung, S.M. & Lee, J.Y. (1998). "Information Discovery on the Internet" . In *Encyclopedia of Library and Information Science*. vol. 62, (supple. 25) pp.146-157.
- Drott, M.C. (2002). "Indexing aids at Corporate websites: the use of robot.txt and meta tags". *Information Processing and Management*, 38(2), 209-219. Retrieved June 10, 2006, From Elsevier database.
- Fadaie, G. (2005). " User Satisfaction through better indexing". *Cataloging & Classification Quarterly*, 40(2), 5-17.
- Fishkin, R. (2005). " Search Engine Ranking Factors" Retrieved Des, 17, 2005, From <http://www.seomoz.org/articles/search-ranking-factors.php#1>
- Global Top 500 (2006) Retrieved Jun, 11, 2006, From http://www.alexa.com/site/ds/top_500
- Google (2004). "Information for webmasters". Retrieved Oct, 11, 2005, From <http://www.google.com/webmasters/4.html>
- Gotlieb, C. ; Eliopoulos, D. (2003). " Evaluating Web Search Results Ranking" . *Online*, 27(2). Retrieved Oct, 11, 2005, From Ebsco Computer Science Database.
- Henshaw, R. & Valauskas, E.J. (2001). "Metadata as a catalyst: experiments with metadata and search engines in the internet journal, First Monday" . *Libri*, 51(2), 86-101pp. Retrieved Oct, 23, 2005, From www.librijournal.org/pdf/1999-3pp125-131.pdf
- Jansen, B.J. ; Spink, A.; and Saracevic, T. (2000). " Real life, real users, and real needs: a study and analysis of user queries on the web" . *Information Processing & Management*, 36(2), 207-227. Retrieve May, 8, 2005, From Elsevier Database.
- Lee-Smeltzer, K.H. (2000). " Finding the needle: Controlled Vocabularies, Resource Discovery, and Dublin core". *Library collections, Acquisitions, & Technical Services*, 24 , 205-215. Retrieve May, 8, 2005, From Elsevier Database.
- Lewis , E. (2005). Top Ten Search Engines - Top 10 SEs . Retrieved Apr, 22, 2006, From <http://www.seoconsultants.com/search-engines/>
- List of Top Search Engines (2006). Retrieved Apr, 22, 2006, From <http://capmex.biz/resources/top-search-engines>

- Noruzi, A. (2005). "Editorial" .*Webology*, 2(4). Retrieved Apr, 20, 2006, From <http://www.webology.ir/2005/v2n4/editorial6.html>
- Quevedo_Torrero, J.U. (2004). "Improving Web Retrieval by Mining the HTML tags for Keywords and Exploring the Hyperlink Structures of Web Pages" [Abstract] doctoral Dissertation. University of Houston. Retrieved Oct, 10, 2005, From <http://wwwlib.umi.com/dissertations/fullcit/3156028>
- Safari, M. (2005). "Search Engines and Resource Discovery on the web: Is Dublin Core an Impact Factor?". *Webology*, 2(2).Retrieved Apr, 20, 2006, From <http://www.webology.ir/2005/v2n2/a13.html>
- Sullivan, D. (2002). "How to Use HTML Meta tags". Retrieved Apr,20, 2006, From <http://searchenginewatch.com/webmasters/atricle.php/2167931>
- Thelwall, M.; Vaughan, L. (2004)." New Versions of PageRank employing alternative web document models" . *ASLIB Proceedings*, 56(1), 24-33. Retrieved Feb, 19, 2005, From www.scit.wlv.ac.uk/~cm1993/papers/2004_new_pagerank_preprint.pdf
- Turner, T.P.; Brackbill, L. (1998). "Rising to the Top: Evaluating the use of html meta tag to improve retrieval of world wide web documents through internet search engines" . *Library Resources and Technical Services*, 24(4), 258-271.
- Yaltaghian , B. (2002). "Improving the ranking of Search Engine Output: A Network Analysis Approach". [Abstract] doctoral Dissertation. University of Toronto. Retrieved Feb, 21, 2006, From www.cosc.canterbury.ac.nz/open/seminars/old/2002.html
- Yu et al (2002). "Patterns in Unstructured Data: Discovery, Aggregation, and Visualization". Retrieved Feb, 21, 2005, From http://javelina.cet.middlebury.edu/lisa/out/cover_page.htm
- Zhang, Z. ; Dimitroff, A. (2005b). " The Impact of Metadata Implementation on Webpage Visibility in Search Engine Results (Part II)" . *Information Processing & Management*, 41 (3), 691-715. Retrieved Feb, 18, 2005, From Elsevier Database.
- Zhang, Z. ; Dimitroff, A. (2004). "Internet Search engines response to metadata Dublin core implementation" . *Journal of Information Science* , 30(4), 310-320. Retrieved Oct, 11, 2005, From Elsevier Database.
- Zhang, Z. ; Dimitroff, A. (2005a). " The Impact of Webpage Content Characteristics on Webpage Visibility in Search Engine Results (Part I)" . *Information Processing & Management*, 41 (3), 665-690. Retrieved Feb, 18, 2005, From Elsevier Database.
- Zhang, Z. ; Jastram, I. (2005). " A Study of the metadata creation behavior of different user groups on the Internet" *Information Processing and management*. Retrieved Oct, 11, 2005, From Elsevier Database.